UNIVERSITÀ DEGLI STUDI DI TERAMO

Dal campione alla popolazione: STIME

puntuali e per intervalli (2 di 2)

Karina Iuvinale – L-GASTR UniTe – a.a. 2024/2025

Intervallo di Confidenza

per la STIMA di una PROPORZIONE

Intervallo di Confidenza per la STIMA di una PROPORZIONE

Quando vogliamo stimare una proporzione (o, in altre parole, la percentuale di unità statistiche che presentano un determinato attributo) è possibile estrarre un campione dalla popolazione di riferimento e calcolare tale proporzione all'interno del campione.

È possibile dimostrare che, se potessimo calcolare tutte le proporzioni all'interno di tutti i campioni di una certa numerosità n estraibili da una determinata popolazione, tali proporzioni danno origine ad una variabile casuale che si distribuisce come una curva normale con valore atteso pari a **π** e varianza pari a **$\pi(1-\pi)/n$**

$$\frac{p-\pi}{\sqrt{\frac{\pi\cdot(1-\pi)}{n}}}$$

per n sufficientemente grande [almeno pari a 5 sia il prodotto $n\pi$, sia il prodotto $n(1-\pi)$] si distribuisce come una V.C. Normale standardizzata

Stima di una proporzione
Esercizio 1

In una grande ditta che si occupa di *food and beverage*, **772 addetti del settore marketing** hanno la facoltà di partecipare ad un corso formativo specialistico. Al fine di stimare la proporzione di coloro che potrebbero aderire all'iniziativa formativa, l'azienda **sorteggia un campione in blocco di 27 addetti**, di cui 17 si dichiarano favorevoli. Fornire una stima della proporzione di **potenziali partecipanti al corso**, a un livello di confidenza del 95%.

Soluzione:

Siano n = 27 il campione estratto e $p = \frac{17}{27} = 0{,}63$ la proporzione campionaria degli addetti favorevoli all'iniziativa tra tutti quelli estratti.

Con un livello di confidenza del 95%, si ha che $1 - \alpha = 0{,}95$, allora $\alpha = 0{,}05$.

Il valore critico (coefficiente) da inserire nell'intervallo sarà allora = 1,96.

L'intervallo di confidenza al 95% per la proporzione della popolazione sarà quindi:

formula che uso per calcolare l'errore in una proporzione

$$IC_{95\%} = \hat{p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} =$$

$$= 0.63 \pm 1.96\sqrt{\frac{0.63 \cdot (1-0.63)}{27}} = 0.63 \pm 0.1821$$

$$IC_{95\%} = (0.63 - 0.1821, 0.63 + 0.1821) = (0.4479, 0.8121)$$

Questo vuol dire che abbiamo una probabilità del 95% che la proporzione della popolazione si trovi all'interno di questo intervallo.

Stima di una proporzione
Esercizio 2

Si vuole conoscere la proporzione di persone che sono state sottoposte a tutte le vaccinazioni previste dalla normativa. A tale scopo, viene effettuata un'indagine su 80 unità statistiche, 60 delle quali risultano in regola. Si determini l'intervallo di confidenza, ad un livello di fiducia del 95%, per la stima della proporzione di persone che sono in regola con le vaccinazioni obbligatorie.

$$p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \pi \le p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$0{,}75 - 1{,}96\sqrt{\frac{0{,}75 \cdot 0{,}25}{80}} \le \pi \le 0{,}75 + 1{,}96\sqrt{\frac{0{,}75 \cdot 0{,}25}{80}}$$

$$0{,}75 - 1{,}96 \cdot 0{,}048 \le \pi \le 0{,}75 + 1{,}96 \cdot 0{,}048$$

$$0{,}75 - 0{,}09 \le \pi \le 0{,}75 + 0{,}09$$

$$\boxed{0{,}66 \le \pi \le 0{,}84}$$

$$n = 80 \quad ; \quad p = \frac{60}{80} = 0{,}75$$

$$(1-p) = 1 - 0{,}75 = 0{,}25$$

$$np = 80 \cdot 0{,}75 = 60$$

$$n(1-p) = 80 \cdot 0{,}25 = 20$$

$$z_{0{,}025} = 1{,}96$$

Stima di una proporzione
Esercizio 3

In un campione di 400 persone alle quali è stato somministrato un vaccino, 136 di esse hanno avuto effetti collaterali di un certo rilievo. Determinare un intervallo di confidenza con un livello di fiducia del 95%, della proporzione della popolazione che soffre di tali effetti collaterali.

$$n = 400$$

$$p = \frac{136}{400} = 0{,}34$$

$$(1 - p) = 1 - 0{,}34 = 0{,}66$$

$$p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \pi \le p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$0{,}34 - 1{,}96 \cdot \sqrt{\frac{0{,}34 \cdot 0{,}66}{400}} \le \pi \le 0{,}34 + 1{,}96 \cdot \sqrt{\frac{0{,}34 \cdot 0{,}66}{400}}$$

$$0{,}34 - 1{,}96 \cdot 0{,}0237 \le \pi \le 0{,}34 + 1{,}96 \cdot 0{,}0237$$

$$0{,}34 - 0{,}046 \le \pi \le 0{,}34 + 0{,}046$$

$$\boxed{0{,}294 \le \pi \le 0{,}386}$$

$$np = 400 \cdot 0{,}34 = 136$$

$$n(1-p) = 400 \cdot 0{,}66 = 264$$

$$(1-\alpha) = 0{,}95 \Rightarrow z_{\alpha/2} = 1{,}96$$

Intervallo di Confidenza
PER PICCOLI CAMPIONI

la t di Student

Intervallo di Confidenza
PER PICCOLI CAMPIONI

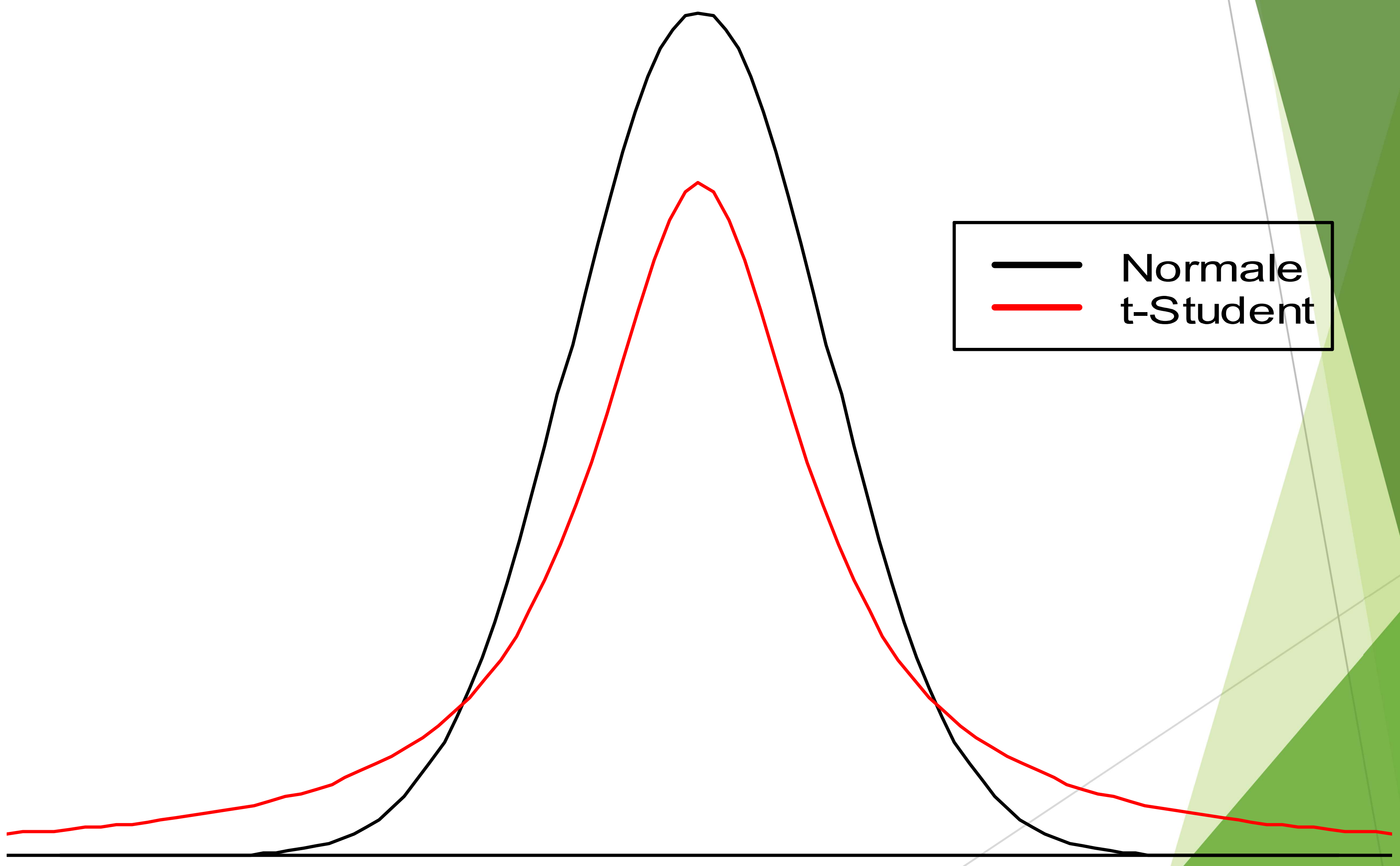
Quando abbiamo a che fare con campioni di bassa numerosità, considerato anche che, nella maggior parte dei casi, la varianza «vera» della popolazione è un parametro sconosciuto, **non sarà possibile fare riferimento alla curva normale come modello generatore del fenomeno** che stiamo considerando (ossia, la stima per intervallo della media aritmetica della popolazione)

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \longrightarrow$$

T-STUDENT
con (n-1) Gradi di Libertà

$$P\left(-t_{\alpha/2} \le \frac{\overline{X}-\mu}{S/\sqrt{n}} \le +t_{\alpha/2}\right) = 1-\alpha \longrightarrow \overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

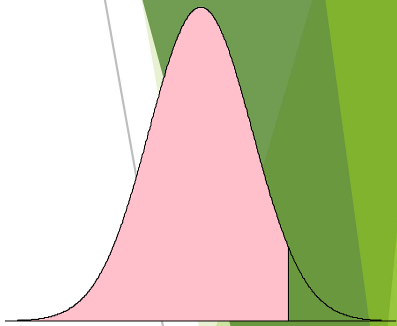
Curva Normale e Curva T-Student



Valori critici della T-Student

con (n-1) Gradi di Libertà

	Area nella coda destra					
Gradi di libertà	**0,250**	**0,100**	**0,050**	**0,025**	**0,010**	**0,005**
1	1,0000	3,0777	6,3137	12,7062	31,8210	63,6559
2	0,8165	1,8856	2,9200	4,3027	6,9645	9,9250
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8408
4	0,7407	1,5332	2,1318	2,7765	3,7469	4,6041
5	0,7267	1,4759	2,0150	2,5706	3,3649	4,0321
6	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,7111	1,4149	1,8946	2,3646	2,9979	3,4995
8	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,6998	1,3722	1,8125	2,2281	2,7638	3,1693
50	0,6794	1,2987	1,6759	2,0086	2,4033	2,6778
70	0,6780	1,2938	1,6669	1,9944	2,3808	2,6479
80	0,6776	1,2922	1,6641	1,9901	2,3739	2,6387
100	0,6770	1,2901	1,6602	1,9840	2,3642	2,6259
110	0,6767	1,2893	1,6588	1,9818	2,3607	2,6213
120	0,6765	1,2886	1,6576	1,9799	2,3578	2,6174
150	0,6761	1,2872	1,6551	1,9759	2,3515	2,6090
200	0,6757	1,2858	1,6525	1,9719	2,3451	2,6006
∞	0,6745	1,2816	1,6449	1,9600	2,3264	2,5758



$$(1-\alpha) = 0{,}95$$
$$\alpha = 0{,}05$$
$$\alpha / 2 = 0{,}025$$

Intervallo di Confidenza per Piccoli Campioni
Esercizio 1

Si vuole conoscere il peso medio dei neonati venuti alla luce in un certo ospedale. Si estrae un campione casuale di 16 elementi, e si riscontra un peso medio di 3,42 kg, con una varianza campionaria pari a 0,4624. Costruire un intervallo di confidenza ad un livello di fiducia del 99% per la stima del peso medio della popolazione di neonati.

$$n = 16 \qquad \bar{x} = 3{,}42 \qquad s^2 = 0{,}4624 \qquad s = \sqrt{s^2} = \sqrt{0{,}4624} = 0{,}68$$

	Area nella coda destra					
Gradi di libertà	**0,25**	**0,10**	**0,05**	**0,025**	**0,01**	**0,005**
1	1,0000	3,0777	6,3137	12,7062	31,8210	63,6559
2	0,8165	1,8856	2,9200	4,3027	6,9645	9,9250
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8408
4	0,7407	1,5332	2,1318	2,7765	3,7469	4,6041
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
12	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545
13	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123
14	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768
15	0,6912	1,3406	1,7531	2,1315	2,6025	2,9467
16	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208
17	0,6892	1,3334	1,7396	2,1098	2,5669	2,8982

$$(1-\alpha) = 0{,}99$$

$$\alpha / 2 = 0{,}005$$

$$g.d.l. = 16 - 1 = 15$$

(n-1) Gradi di Libertà

$$n = 16 \qquad \bar{x} = 3{,}42 \qquad s^2 = 0{,}4624 \qquad s = \sqrt{s^2} = \sqrt{0{,}4624} = 0{,}68$$

$$(1-\alpha) = 0{,}99 \Rightarrow t_{\alpha/2;15} = 2{,}947$$

$$\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$3{,}42 - 2{,}947 \cdot \frac{0{,}68}{\sqrt{16}} \le \mu \le 3{,}42 + 2{,}947 \cdot \frac{0{,}68}{\sqrt{16}}$$

$$3{,}42 - 0{,}5 \le \mu \le 3{,}42 + 0{,}5$$

$$\boxed{2{,}92 \le \mu \le 3{,}92}$$

Esercizio 2

Un'azienda farmaceutica vuole stimare la quantità di principio attivo presente in un certo medicinale. Viene estratto un campione casuale di 25 flaconi del farmaco, e si trova una quantità media di principio attivo nel campione pari a 20,8 mg, con una varianza pari a 1,44. Costruire un intervallo di confidenza ad un livello di fiducia del 95% per la stima della quantità di principio attivo presente nel medicinale.

	Area nella coda destra					
Gradi di libertà	**0,25**	**0,10**	**0,05**	**0,025**	**0,01**	**0,005**
1	1,0000	3,0777	6,3137	12,7062	31,8210	63,6559
2	0,8165	1,8856	2,9200	4,3027	6,9645	9,9250
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8408
4	0,7407	1,5332	2,1318	2,7765	3,7469	4,6041
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
21	0,6864	1,3232	1,7207	2,0796	2,5176	2,8314
22	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188
23	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073
24	0,6848	1,3178	1,7109	2,0639	2,4922	2,7970
25	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874
26	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787

$n = 25$

$\bar{x} = 20,8$

$s^2 = 1,44$

$s = \sqrt{s^2} = 1,2$

$(1-\alpha) = 0,95$

$\alpha = 0,05$

$\alpha / 2 = 0,025$

$g.d.l. = 25 - 1 = 24$

(n-1) Gradi di Libertà

$$\overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$(1-\alpha) = 0{,}95 \Rightarrow t_{\alpha/2;24} = 2{,}0639$$

$$20{,}8 - 2{,}0639 \cdot \frac{1{,}2}{\sqrt{25}} \leq \mu \leq 20{,}8 + 2{,}0639 \cdot \frac{1{,}2}{\sqrt{25}}$$

$$20{,}8 - 2{,}0639 \cdot 0{,}24 \leq \mu \leq 20{,}8 + 2{,}0639 \cdot 0{,}24$$

$$20{,}8 - 0{,}49 \leq \mu \leq 20{,}8 + 0{,}49$$

$$\boxed{20{,}31 \leq \mu \leq 21{,}29}$$

$$n = 25$$
$$\overline{x} = 20{,}8$$
$$s^2 = 1{,}44$$
$$s = \sqrt{s^2} = 1{,}2$$

Determinazione della numerosità campionaria

Come si determina la numerosità campionaria per la stima di una media o di una proporzione? Cenni…

Determinazione della numerosità campionaria per la stima della media

È necessario ricordare la differenza tra piccolo e grande campione. Devo però considerare che, in alcuni casi, si può stabilire una numerosità campionaria minima per effettuare una stima (come negli esempi che seguono).

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad\qquad \varepsilon = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Moltiplicando entrambi i membri per la quantità: $\sqrt{n}$ $\longrightarrow$ $\varepsilon\sqrt{n} = z_{\alpha/2} \cdot \sigma$

Dividendo entrambi i membri per la quantità: ε $\longrightarrow$ $\sqrt{n} = \dfrac{z_{\alpha/2} \cdot \sigma}{\varepsilon}$

$$\boxed{n = \frac{z^2_{\alpha/2}\,\sigma^2}{\varepsilon^2}}$$

Determinazione della numerosità campionaria per la stima della media (2)

L'età degli studenti di una certa Facoltà si distribuisce come la v.c. normale, con varianza pari a 45. Quale numerosità campionaria minima è necessaria per stimare un intervallo di confidenza dell'età della popolazione, utilizzando un livello di fiducia pari al 95% ed accettando un errore massimo ammissibile pari a 0,7 anni?

$$n = \frac{z_{\alpha/2}^2 \cdot \sigma^2}{\varepsilon^2} = \frac{(1{,}96)^2 \cdot 45}{(0{,}7)^2} = \frac{3{,}8416 \cdot 45}{0{,}49} = 352{,}8$$

Determinazione della numerosità campionaria per la stima della proporzione

$$n = \frac{z^2_{\alpha/2}\,\sigma^2}{\varepsilon^2} \longrightarrow \boxed{n = \frac{z^2_{\alpha/2}\cdot\pi(1-\pi)}{\varepsilon^2}}$$

Si vuole stimare la percentuale di individui affetti da una certa patologia. Determinare la numerosità campionaria minima necessaria affinché la proporzione della popolazione cada in un intervallo di livello di fiducia pari al 90% ed accettando un errore massimo ammissibile pari al 4%

$$n = \frac{z^2_{\alpha/2}\cdot\pi(1-\pi)}{\varepsilon^2} = \frac{(1{,}65)^2\cdot 0{,}5\cdot(1-0{,}5)}{(0{,}04)^2} = \frac{2{,}72\cdot 0{,}25}{0{,}0016} = 425$$

Mini-Glossario

STIMA PUNTUALE DELLA MEDIA

è data dalla media aritmetica delle modalità presentate dalle unità statistiche facenti parte di un campione; in sostanza, è un singolo "numero" calcolato sui dati campionari, e viene utilizzato per stimare un parametro (la media aritmetica, ad esempio) della popolazione.

STIMA PER INTERVALLO DELLA MEDIA

è un intervallo di valori all'interno del quale si presume sia compreso il valore del parametro (ignoto) della popolazione.

INTERVALLO DI CONFIDENZA

è un intervallo di valori al quale è associata una probabilità (per essere precisi, un «livello di confidenza») indicata, generalmente, con $(1-\alpha)$. Tale probabilità quantifica la possibilità che all'interno dell'intervallo sia compreso il vero parametro della popolazione.

LIVELLO DI CONFIDENZA (O DI FIDUCIA) – *Confidence Level*

è dato dal complemento ad uno di (1-α), dove α rappresenta la massa di probabilità che si trova nelle code della distribuzione, al di fuori dell’intervallo di confidenza.

Rappresenta, in un certo senso, una misura dell’«accuratezza» con la quale è costruito l’intervallo di confidenza; maggiore è il livello di fiducia, maggiore sarà la probabilità che il parametro che vogliamo stimare sia compreso all’interno di tale intervallo.

GRANDI CAMPIONI E PICCOLI CAMPIONI

per grandi campioni intendiamo campioni che hanno una numerosità almeno pari a 60-70 elementi (tale soglia è puramente indicativa: consultando differenti testi, potreste trovare differenti soglie)

Riferimenti sul testo

di **Whitlock M.C., Schluter D.**
***Analisi statistica dei dati biologici*, Zanichelli**

Paragrafi da studiare: 4.1, 4.2., 4.3, 7.3 (tranne il metodo di Wald). Esercizi alla fine dei paragrafi.

di **Triola M. M., Triola M. F.**
***Statistica per le discipline biosanitarie*, Pearson-Addison Wesley**

Paragrafi da studiare: 6.1, 6.2, 6.3, 6.4. Esercizi alla fine dei paragrafi.